# The effect of informational load on disfluencies in interpreting

## A corpus-based regression analysis

Koen Plevoets and Bart Defrancq

This article attempts to measure the cognitive or informational load in interpreting by modelling the occurrence rate of the speech disfluency *uh(m)*. In a corpus of 107 interpreted and 240 non-interpreted texts, informational load is operationalized in terms of four measures: delivery rate, lexical density, percentage of numerals, and average sentence length. The occurrence rate of the indicated speech disfluency was modelled using a rate model. Interpreted texts are analyzed based on the interpreter's output and compared with the input of non-interpreted texts, and measure the effect of source text features. The results demonstrate that interpreters produce significantly more *uh(m)*s than non-interpreters and that this difference is mainly due to the effect of lexical density on the output side. The main source predictor of *uh(m)*s in the target text was shown to be the delivery rate of the source text. On a more general level of significance, the second analysis also revealed an increasing effect of the numerals in the source texts and a decreasing effect of the numerals in the target texts.

**Keywords:** interpreting, cognitive load, disfluencies, rate model, corpus linguistics

## 1. Introduction

Cognitive demand during the interpreting task arises from the division of attention between listening to the input utterance and orally rendering the translation. With multiple tasks simultaneously demanding cognitive resources, interpreters can be said to operate constantly under heavy cognitive pressure. In light of this high processing load, Gile (1999) formulated the 'tightrope hypothesis,' stating that, all things being equal, interpreters normally work at the limits of their processing capacities.

*Translation and Interpreting Studies* 11:2 (2016), **202–224**. DOI 10.1075/tis.11.2.04ple
ISSN 1932–2798 / E-ISSN 1876–2700 

Since initial inquiries into interpreting, researchers have attempted to represent the various stages of information processing during the interpreting task using models. For instance, Gerver (1975) and Moser-Mercer (1978) present models for processing, which subsequently were outlined by Paradis (1994) from a neurolinguistic perspective and by Setton (1999: 65) from a pragmatic-linguistic perspective. Gile's Effort Model (Gile 1997, 2009) and Seeber's (2011) Cognitive Load Model explicitly represent the interpreting process as a 'cognitive management problem.' These models consider interpreting to be the maintenance of an equilibrium between various cognitive demands centered around three challenges: a language comprehension task, a language production task, and memory storage. Gile's Effort Model additionally includes a coordination task.

Due to their focus on the 'processing costs' of interpreting, these models also have explanatory power in that they aim to predict when translation goes awry. Information overflow — i.e., when the processing load exceeds that of the interpreter's processing capacity — is one instance in which the oral rendition is less than ideal. Such information overflow is traditionally held responsible for the so-called errors and omissions: when an interpreter runs into difficulty processing a certain input segment, this may result in an inaccurate translation or a failure to render the segment altogether. Dillinger (1994) and Tommola and Helevä (1998) present corroborating evidence in studies that demonstrate that translations become less accurate when the source texts are propositionally denser.

However, the issue of what constitutes an 'interpreting error' is not without problems. Barik (1975) set out to develop a typology of 'translation departures' in interpreting, distinguishing between the general categories of *omissions*, *additions,* and *substitutions and errors*, with each category being further divided into subtypes. Barik's typology, however, was heavily criticized by Gerver (1976) for being too subjective to apply. Since then, the precise definition of an 'interpreting error' has been wanting.

As a consequence, interpreting researchers shifted their attention to *disfluencies*, such as false starts or repairs, and silent or filled pauses, the latter of which is most typically exemplified by *uh(m)* (Levelt 1983). Disfluencies are generally regarded by psycholinguists as a means of indirectly examining cognitive load, and experiments suggest their use to be symptomatic of new information (Arnold et al. 2003; Clark and Fox Tree 2002) or heavy constituents (Arnold et al. 2000; Watanabe et al. 2008; Swerts 1998).[1]

Ever since Goldman-Eisler's (1967) inauguration of 'pausological' studies in interpreting, most research has been done on silent and filled pauses. Tissi (2000), for instance, analyzed a corpus of ten students' interpretations of two excerpts

1. For an overview of research in this area, see Bortfeld et al. (2001).

from political roundtables. The study found that silent pauses are fewer but longer in target texts than in source texts, but that filled pauses exhibit considerable individual variation. Cecot (2001) replicated this finding in a study with eleven professional interpreters. Finally, Mead (2000) demonstrated that fifteen students produced more filled pauses when interpreting into their B language than into their A language (the results were not significant for the silent pauses). All these findings lend support to a scheme such as summarized by Setton (1999: 247), reproduced in Table 1.

**Table 1.** Disfluencies and attention, from Setton (1999: 247)

| | Attention to input | Attention to formulation |
|---|---|---|
| Long silent pause | High | – |
| Short pausing | Normal listening | Routine planning |
| Filled pause | Normal listening | Routine planning |
| Mixed: short & filled pauses and voice effects | Normal listening | Routine planning |
| Long filled pause | Relaxed or off | Planning/Searching |
| Fluent unmodulated string | Relaxed of off | Off |

However, the occurrence of disfluencies in interpreting has never been systematically investigated in naturalistic interpreting settings. Unlike other studies, this study will take both a classic and a 'comparable' approach. The classic approach studies the influence of the input (the source text) on the output (the target text). This is the approach taken by Gerver (1969), Barik (1975), and Shlesinger (2003). In the terminology of Gile's Effort Model, the classic approach can be equated to the study of the listening effort, i.e. the amount of cognitive load required in order to process the critical features of the source text. The comparable approach, on the other hand, contrasts data from interpretations with data from spontaneous non-interpreted speech. This approach has become customary in the field of translation studies since Baker (1993) and can be used to study which features distinguish translations from non-translated text. In our study, the comparison of spontaneous speech and interpreting will be used to measure the effects of the production effort on the frequency of filled pauses (i.e., *uh(m)*'s). Interpreters experience cognitive load not only on the input side but also on the output side, as they have to formulate their translations within a very limited time frame. Comparing the features of their utterances with the features of spontaneous speech, both in relation with disfluencies, will enable us to identify what aspects of the production effort cause disfluencies.

This study will thus be the first to cover the disfluency-inducing aspects of both source text and interpretation. Previous studies have focused solely on source text features and the particular patterns of interpreter behavior which they elicit. It is relevant and important to distinguish both aspects as the efforts involved are different. For instance, numbers may be cognitively more demanding to process than to produce, especially for interpreters (Mazza 2001; Pinochi 2009). Interpreters also apply various strategies to reduce cognitive load both on the input side (e.g., shallow parsing, cognate translation; see Riccardi 1998) and on the output side (e.g., chunking of complex sentences, reordering of information, anticipation). Consequently, source texts and interpretations exhibit different features of cognitive load, and information overload is expected to be more of a problem on the input side than on the output side: interpreters do not have any effect on what they hear, but they do control what they say. Our study will enable us to tell whether these expectations are fulfilled.

In short, the research questions we seek to answer are:

1. What aspects of cognitive load lead to disfluencies for both spontaneous speakers and interpreters?
2. Do interpreters struggle more with cognitive load on the input side than on the output side?

In the next sections, we will successively define and operationalize the concept of cognitive load and describe the data. Then, the corpus-based statistical methodology employed to answer the research questions will be discussed. Results of this analysis are then presented and discussed within the context of the previously mentioned models with some general conclusions drawn at the end of the article.

## 2. Cognitive load: Definition and causes

Cognitive load, information overload, and related concepts all rest on the idea, first explored in the 1950s by Welford (1952) and Broadbent (1958), that the human working memory has only limited capacity, which prevents it from performing several tasks simultaneously at the same speed and the same level of efficiency as when the tasks are performed separately. It has also limited capacity for storing the information that is necessary to perform the tasks. Over the years, models of information processing in humans have evolved, but the overall picture of matching demands with capacities is still very much in use. In interpreting, Gile (1997) and Seeber (2011) are examples of the capacity-demand model.

Cognitive load is defined by Seeber (2011: 187) as “the amount of load generated by individual concurrent tasks” involved in a cognitive process, where load is represented by the demands of the individual tasks. Load is determined both at the macrostructural level, i.e., the level of the major task categories (perceptual, cognitive, response), and at the microstructural level, i.e., the level of linguistic chunks. Cognitive load not only depends on the resources required by the individual tasks but also by the extent to which they interfere with each other. Tasks that involve similarly-structured processing dimensions interfere more than tasks relying on different structures, thus increasing the cognitive load.

Seeber (2011) lists two input features that are known to increase the cognitive load in simultaneous interpreting: delivery rate and late verb placement. On the output side, Seeber describes several interpreter strategies for dealing with differences in verb placement between source and target language in terms of the cognitive load they generate. However, the article does not provide empirical evidence in support of these theoretical constructs. Seeber (2013) discusses the potential of pupillometry to generate empirical evidence to investigate cognitive load, but does not include a dataset.

Our study will use the frequency of filled pauses or *uh(m)*'s as evidence of increased cognitive load, in reference to the findings reported in the psycholinguistic literature (cf. the studies mentioned in the previous section). The independent variables that will be analyzed as having a potential effect on cognitive load include: source text delivery rate, the proportion of numbers, the lexical density, and the average sentence length. In this study, each of these measures will be computed for each text as a whole.

Delivery rate is reported not to increase cognitive load significantly in the comprehension of spontaneous speech (Voor and Miller 1965) but has been shown to considerably influence interpreter performance (Gerver 1969; Pio 2003). The same can be said about numbers: to our knowledge, there is no evidence in the cognitive literature suggesting that the comprehension or production of numbers in spontaneous speech increases the cognitive load in the hearer or in the speaker respectively. There is, however, ample evidence that the presence of numbers affects the interpreter's performance (Gile 2009) and constitutes one of the most important sources of stress in the interpreting community (Alessandrini 1990).

Lexical density and (average) sentence length are traditional measures of text difficulty. Lexically dense texts are harder for readers to comprehend and their retention scores are generally — but not always — worse, indicating that the subjects experience difficulties in coping with the cognitive load generated by a high lexical density (Kintsch et al. 1975; Gibson 1993). Gile (2008) points to the lexical density as one of the prime determinants in cognitive load in interpreting. Sentence length, in contrast, is a traditional readability parameter (Flesch 1948) because long sentences are assumed to be harder for readers to process and to retain. In interpreting, Gile (2008) highlights source text sentence length as one of the factors increasing cognitive load in interpreters, adding that length as such is probably not a factor, but the syntactic complexity that comes with it in most cases. Chmiel and Mazur (2013) report that in an experiment on sight translation performed by trainee interpreters long sentences receive longer fixation times, indicating an increased cognitive load with the interpreter.

Research into the factors of cognitive load in interpreting has been predominantly source-oriented. The features of the target text are usually not deemed relevant enough to be included in the cognitive load of the interpreter. However, it is reasonable to assume that the cognitive load of an interpreter is also influenced by his or her attempts to *produce* a target text at high speed, with high lexical density, a high proportion of numbers, and with long sentences. As the aim of this study is to examine the cognitive load on both the input and the output side, we will also measure the four parameters of cognitive load in the target texts.

## 3. Data

Corpus-based studies are still relatively rare in interpreting studies. Shlesinger (1998) was the first to call upon the research community to compile corpora with interpreting data and only some research institutes have followed suit: Bologna-Forli; Hamburg; Ghent and Poznan (for an overview of current research in corpus-based interpreting studies, see Straniero and Falbo 2012). Corpus data have considerable advantages and can prove useful to investigate cognitive load in more naturalistic settings. Since the data are collected from real-life interpreting events, these data can overcome some of the challenges traditionally associated with highly-controlled experimental settings. Corpora are typically designed to examine a variety of linguistic phenomena and paralinguistic conditions that are representative of language in use. Consequently, they should cover as many instances as possible of the linguistic activity under scrutiny and not just one or two phenomena, as is typically the case in research based on ad hoc recordings of an individual interpreter at a particular conference. In sum, corpus data have the advantage of representing actual instances of interpreting carried out by a variety of professionals under working conditions.

The data used for this study are collected from a corpus of interpreted Dutch and from a corpus of non-interpreted Dutch. The corpus of interpretations was compiled at the Department of Translation, Interpreting, and Communication of Ghent University. It consist of plenary speeches and their interpretations recorded in the European Parliament from 2006 until 2008. The source languages in the corpus are currently restricted to French, Spanish, English, and Dutch. The available target languages are Dutch, French, and English. The audio-visual fragments (which are downloadable on the website of the European Parliament) are transcribed according to the guidelines of the VALIBEL corpus (Bachy et al. 2007).The corpus is still under development, and its current size is about 220,000 tokens. For the purpose of this study, only the sub-corpus of the French source speeches and their Dutch interpretations will be used. That sub-corpus has additionally been tagged for parts-of-speech, lemmas, and chunks by means of the 'LeTs Preprocess Toolkit' (Van de Kauter et al. 2013).

The corpus of non-interpreted Dutch is the sub-corpus of parliamentary debates ('component g') of the Spoken Dutch Corpus (Oostdijk 2000). The compilation of the Spoken Dutch Corpus was a joint venture between several Netherlandic and Flemish universities and was undertaken from 1998 until 2003. The Netherlandic part of component g contains about 220,000 tokens and the Flemish part contains about 140,000 tokens. The whole corpus is annotated for parts-of-speech and lemmas.

Table 2 presents some summary frequencies of both corpora. For obvious reasons, the French sub-corpus of EPICG has the same number of files as the Dutch sub-corpus, as the latter are the interpretations of the former. It can be seen that the Spoken Dutch Corpus is somewhat larger, and the Netherlandic texts in particular are also slightly longer. The distinction between the Netherlandic part and the Flemish part will not be taken into consideration in the analysis.

**Table 2.** Summary overview of corpora

| | | No. of files | No. of sentences |
|---|---|---|---|
| EPICG | FRA (source) | 107 | 1455 |
| | DUT (target) | 107 | 1431 |
| | | | |
| SPCg | TOTAL | 240 | 19046 |
| | Flanders | 155 | 8293 |
| | Netherlands | 85 | 10753 |

## 4. Method

In both corpora, the number of *uh*'s and *uhm*'s (or *euh*'s and *euhm*'s, as these are the transcription conventions in the EPICG) were counted by means of a Python script. The *uhm*'s were counted separately from the *uh*'s, but because their frequencies proved to be rather low, they will be taken together with the *uh*'s in the analysis.

The next step consisted of operationalizing the notion of cognitive load in terms of the four 'informational measures' mentioned in Section 2. The delivery rate of a text was defined to be the total number of words of a text, divided by time that the utterance of the text lasted, measured in minutes (i.e., each text contained information on its time lapse in seconds, the total of which was divided by 60). The lexical density of a text was defined as the total number of content words in the text divided by its total number of function words. That distinction was determined on the basis of the part-of-speech tags in the corpora. The content words were taken to be the nouns, adjectives, adverbs derived from adjectives, and non-auxiliary verbs. The function words are the articles, conjunctions, prepositions, pronouns, the so-called pronominal adverbs (e.g., *daarin* 'therein,' *hiernaar* 'hereto,' or *waarvan* 'whereof,' in Haeseryn et al. 1997: 490–503), and auxiliary verbs (initially, the conjunctions and conjunctive adverbs were counted separately in view of Shlesinger's (1995) and Mizuno's (1999) claims about the interpreters' awareness of cohesive markers, but preliminary analyses revealed this measure not to be significant). The third measure of cognitive load is the proportion of numbers in a text. This measure was defined as the total number of numerals in a text (again determined by the part-of-speech tags) divided by the total number of words of the text. Finally, the baseline informational measure of the average sentence length of a text was defined as the total number of words in a text divided by the total number of sentences.

Those four informational measures were used to predict the occurrence rate of *uh(m)* in two analyses. The first analysis was a direct application of the Bakerian 'comparable' approach in that the scores of the 240 non-interpreted texts were compared with the scores of the 107 Dutch target texts. The binary distinction between interpreted Dutch (*NED_in*) and non-interpreted or original Dutch (*NED_or*) was defined as a categorical variable, so that the interactions between this variable and the four informational measures would reveal the different impact of each informational measure in interpreting and spontaneous speech. As mentioned in Section 1, it can be said that because this analysis focuses on the output, its objective is to reveal the difference in production effort between interpreting and non-interpreting. Because the source texts were absent from this analysis, they were taken into account in the second analysis: the occurrence rate of *uh(m)* in the 107 target texts was predicted on the basis of the four informational measures in both the source texts and the target texts. The proportion of *uh(m)* in the 107 source texts was added as the ninth predictor, in order to examine whether disfluencies in the source text facilitate interpreting or not.[2] This second analysis measures the cognitive load of both the source texts and the target texts and aims to capture the listening effort next to the production effort in interpreting.

**2.** The issue of source text disfluencies has been examined by Goldman-Eisler (1967) and Gerver (1975), and to date, there is little consensus. One rationale for its potential usefulness is that these serve as placeholders for the interpreter to fit in as much of the interpretation as possible.

In both analyses, the frequency of *uh(m)* given the text length (as expressed by the total number of words) was modeled as a function of the informational measures. The technical term for such an analysis is a rate model, as it models the relative count of *uh(m)* conditional on the text's total number of words. The total number of words figures as a controlling variable or so-called 'offset' which is itself not estimated but which influences the estimation of the parameters of interest. The details of rate models are extensively discussed in Agresti (2013: 385–391) and programming commands for the R software can be found in Faraway (2006: 61–63).

Table 3 presents some summary statistics on the informational measures. Since all informational measures are somewhat skewed to the right (with the sole exception of the delivery rate of non-interpreted Dutch texts), they were log-transformed (after a normalizing constant of 0.5 was added). The subsequent preliminary analysis furthermore revealed some multicollinearity among the log-transformed measures (as indicated by high Variance Inflation Factors, see Fox

**Table 3.** Summary statistics for the informational measures

| **Delivery rate** | **mean** | **SD** | **skewness** | **kurtosis** |
|---|---|---|---|---|
| NED_or | 157.883 | 23.483 | −0.167 | −0.331 |
| NED_in | 166.053 | 38.855 | 3.865 | 25.211 |
| FRA | 186.718 | 30.338 | 0.584 | 1.339 |
| Total | 166.605 | 31.551 | 1.971 | 14.304 |
| **Lexical density** | **mean** | **SD** | **skewness** | **kurtosis** |
| NED_or | 0.656 | 0.092 | 1.789 | 6.849 |
| NED_in | 0.731 | 0.104 | 0.226 | 0.981 |
| FRA | 0.767 | 0.126 | 0.360 | 0.816 |
| Total | 0.700 | 0.114 | 0.939 | 1.734 |
| **% of Numerals** | **mean** | **SD** | **skewness** | **kurtosis** |
| NED_or | 0.016 | 0.012 | 2.776 | 14.241 |
| NED_in | 0.013 | 0.014 | 1.694 | 3.655 |
| FRA | 0.013 | 0.013 | 1.167 | 0.645 |
| Total | 0.015 | 0.013 | 1.944 | 7.003 |
| **Average Sentence Length** | **mean** | **SD** | **skewness** | **kurtosis** |
| NED_or | 18.093 | 4.777 | 1.015 | 1.943 |
| NED_in | 24.578 | 7.587 | 0.423 | 0.185 |
| FRA | 27.521 | 8.895 | 0.459 | −0.107 |
| Total | 21.843 | 7.801 | 1.027 | 1.023 |

and Weisberg 2011: 325–327). The correlation matrices of the variables in both analyses are shown in the Appendix. Finally, the original rate models, which assumed an underlying Poisson distribution for the frequency of *uh(m)*, exhibited 'overdispersion', i.e., the variance of the counts was disproportionally high in comparison to the mean. The estimated dispersion parameter of each analysis will be given with all other results in the next section. Both rate models were refitted by means of the quasi-Poisson distribution (Faraway 2006: 59–61).

## 5. Analysis

### 5.1 Analysis 1: Comparison of interpreted and spontaneous speech

The rate model of all 347 texts fits well ($G^2 = 2016.640$; $df = 9$; $p < .001$). The estimated dispersion parameter is 12.78, which is taken into account in adjusting the standard errors and residuals of the model. The residuals exhibit only 4.61% (i.e., 16 out of 347 observations) to be greater than two in absolute value. There are no influential observations or so-called leverage points on the estimates (as determined by Cook's D and dfbeta). Table 4 lists the effects of this model, which are also visualized in the 'effect displays' in Figures 1–9. These effect displays are made with the R package *effects* (Fox 2003), and they are highly useful as they automatically transform the effects back to the original log-transformed and standardized variables.

There is a clear difference between non-interpreted Dutch (*NED_or*) and interpreted Dutch (*NED_in*), as the interpreters produce significantly more *uh(m)*'s than non-interpreters (Figure 1). The effect of delivery rate conforms to expectation, as a higher delivery rate leads to an increase of *uh(m)*'s (Figure 2).

**Table 4.** Parameter estimates for Analysis 1

| | Estimate | Std. Error | *t*-value | Pr(>\|t\|) |
|---|---|---|---|---|
| Intercept | −3.791 | 0.054 | −70.608 | <0.001 |
| NED_in | 0.728 | 0.149 | 4.879 | <0.001 |
| Delivery rate | 0.196 | 0.059 | 3.309 | 0.001 |
| Lexical density | −0.321 | 0.070 | −4.555 | <0.001 |
| % of Numerals | −0.159 | 0.065 | −2.430 | 0.016 |
| Avg. sentence length | 0.173 | 0.047 | 3.685 | <0.001 |
| NED_in: Delivery rate | −0.171 | 0.114 | −1.502 | 0.134 |
| NED_in: Lexical density | 0.379 | 0.128 | 2.961 | 0.003 |
| NED_in: % of Numerals | 0.079 | 0.117 | 0.678 | 0.498 |
| NED_in: Avg. Sentence length | −0.135 | 0.108 | −1.257 | 0.210 |

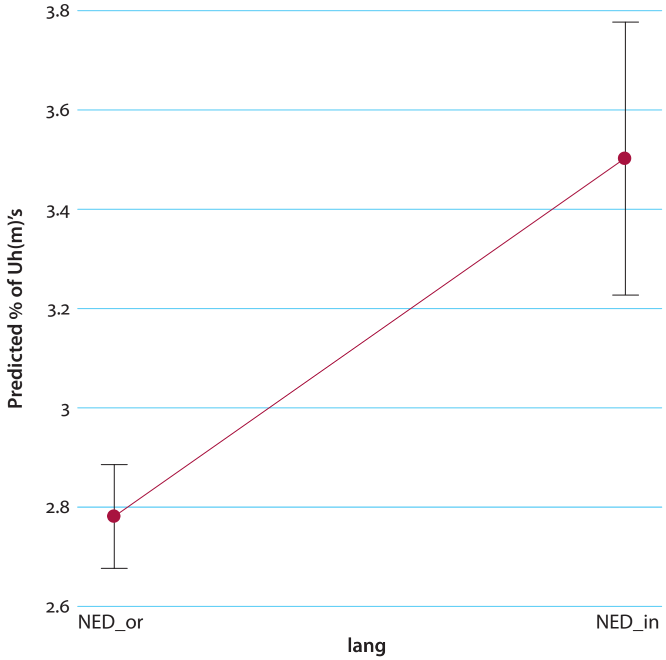


**Figure 1.** Main effect of 'language' (i.e., original vs. interpreted Dutch)

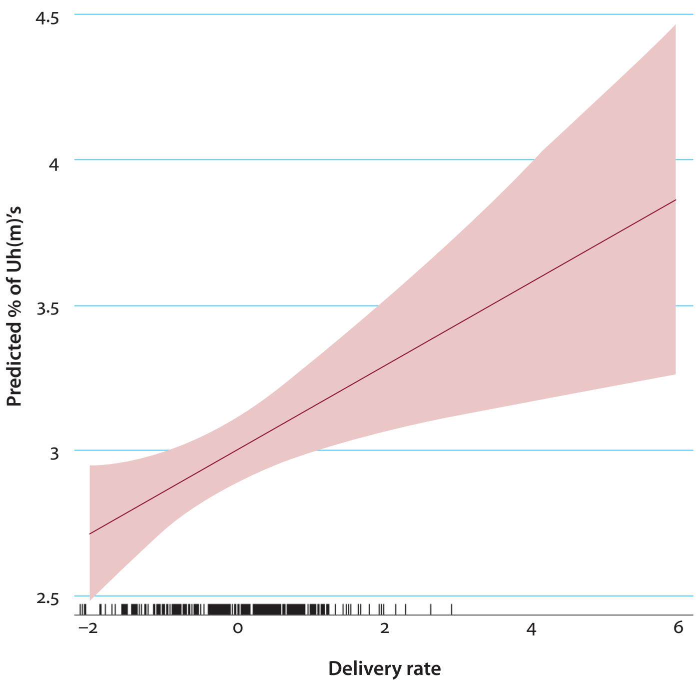


**Figure 2.** Main effect of delivery rate

For interpreting, however, the effect is smaller, as demonstrated by the slope in Figure 3. The difference between the interpreters and the non-interpreters for delivery rate is not statistically significant. That may reflect the fact that different interpreters employ different coping strategies if the delivery rate becomes higher. Future qualitative analyses will have to shed light on this interpretation.

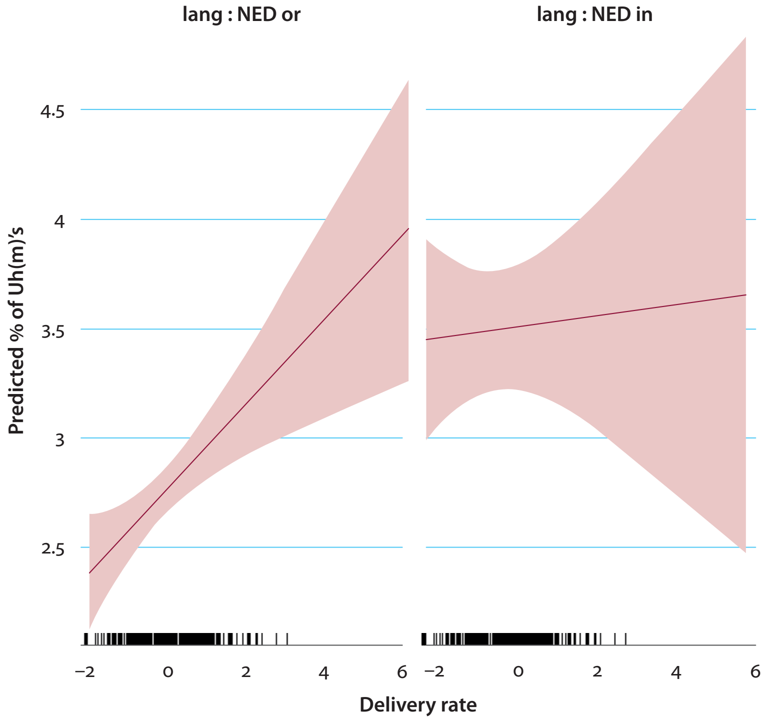


**Figure 3.** Interaction between delivery rate and 'language'

Lexical density shows a decreasing trend for the non-interpreters (Figure 4 and 5). That effect can be attributed to the scripted and prepared nature of the parliamentary speeches, which implies that lexically denser texts indeed result in fewer *uh(m)*'s, when the members of parliament stick more closely to their text. In contrast, the interpreters do show the expected result, which is also statistically significant: interpreters produce more *uh(m)*'s when their interpretations become lexically denser.

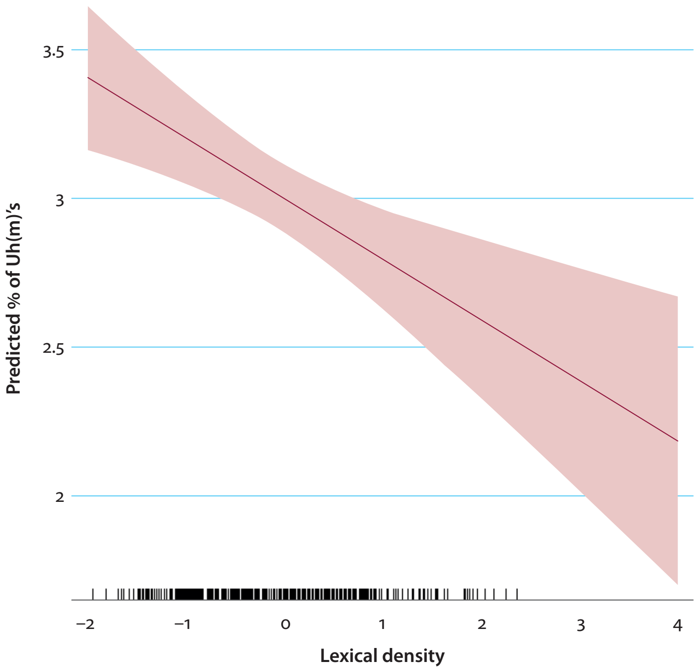


**Figure 4.** Main effect of lexical density

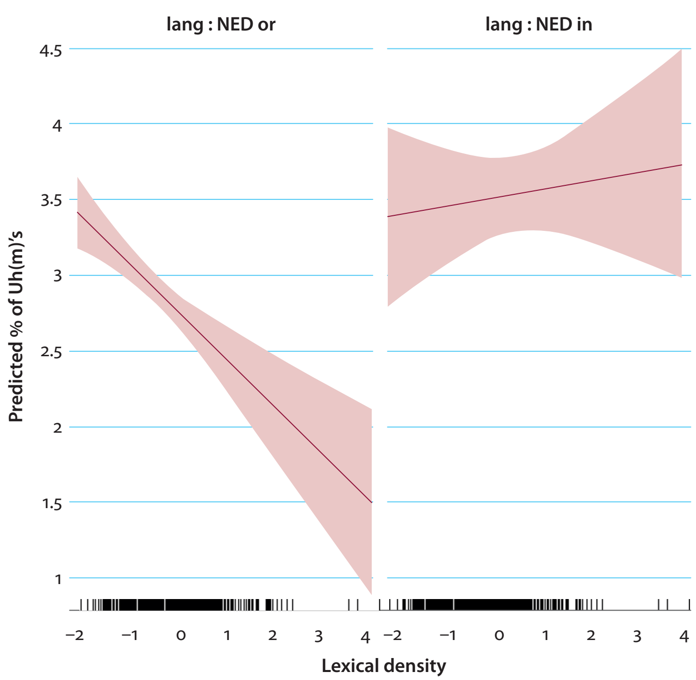


**Figure 5.** Interaction between lexical density and 'language'

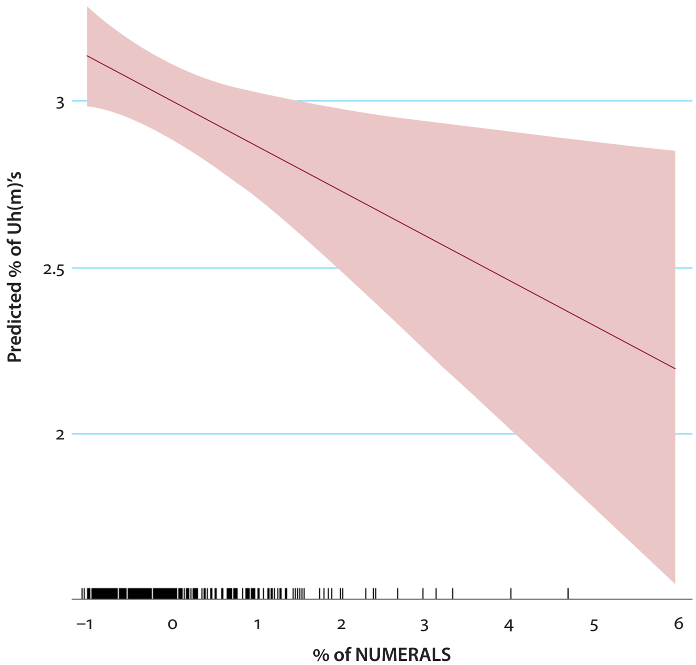


**Figure 6.** Main effect of percentage of numerals

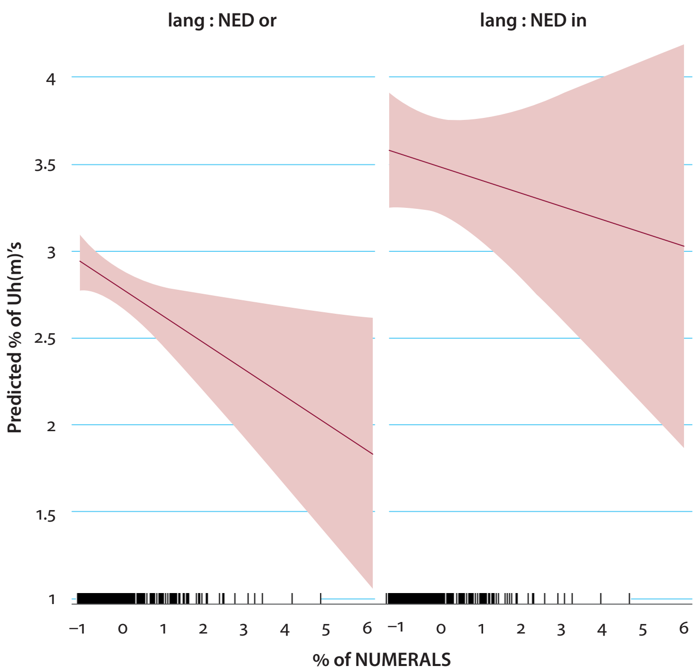


**Figure 7.** Interaction between percentage of numerals and 'language'

The proportion of numerals in a text shows a decreasing effect for both the non-interpreters and the interpreters (Figure 6 and 7). The difference between the two groups is not statistically significant. Nevertheless, it remains clear that interpreters have more difficulties with numbers than non-interpreters, as is evidenced by the fact that the overall effect of the interpreters is higher up the vertical axis than the effect of the non-interpreters (Figure 7). Regardless of any relationship with the proportion of numbers throughout a text, that difference shows that interpreters produce more *uh(m)*'s than non-interpreters when dealing with numbers.

Finally, the average sentence length per text also shows an increasing effect according to expectation (Figure 8 and 9). . The difference between both groups is not statistically significant.

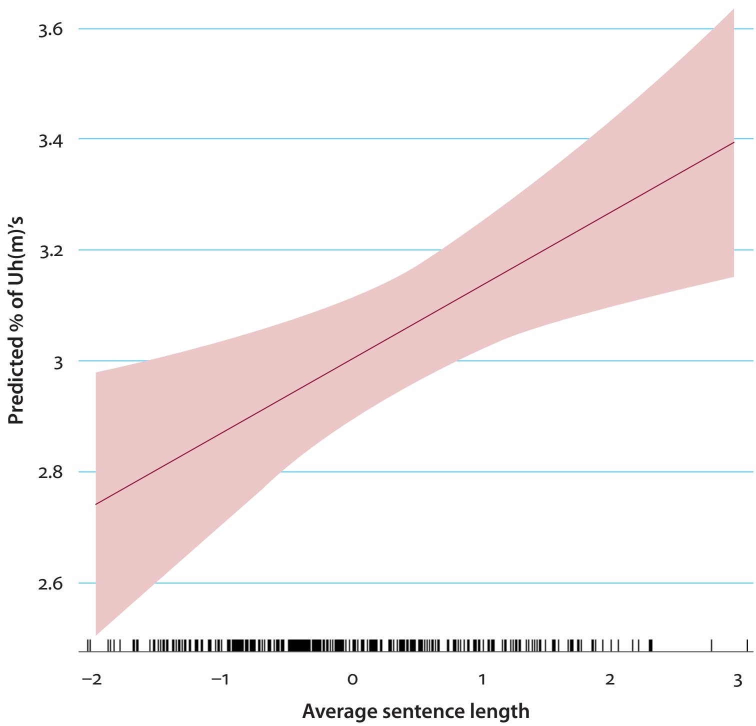


**Figure 8.** Main effect of average sentence length

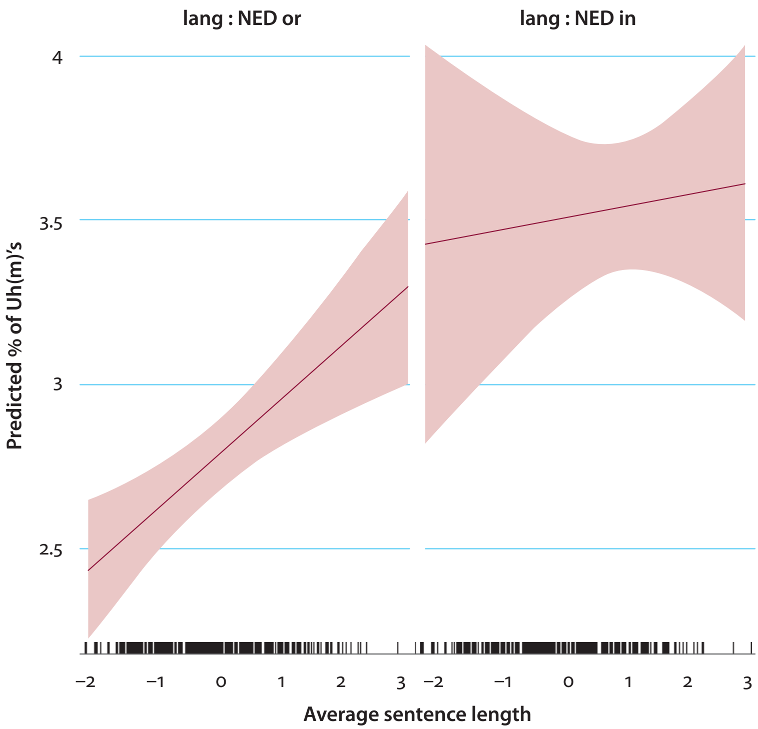


**Figure 9.** Interaction between average sentence length and 'language'

## 5.2 Analysis 2: Comparison of source text and target text

The rate model of the 107 interpretations also fits well ($G^2 = 77.079$, $df = 9$, $p < .001$) with an estimated dispersion parameter of 5.470. There are only 6 residuals outside of the interval [−2,+2], and there are again no leverage points. Table 5 provides the estimated effects (which are informative enough by themselves, so no figures will be given).

It is clear that there are not many significant effects for the rate of *uh(m)*'s in interpreting. Only one predictor stands out, the delivery rate of the source text, which also shows a positive trend as expected: if the delivery rate of the source text increases, the interpreter produces significantly more *uh(m)*'s. That result was also obtained in an analysis with just the five predictors of the source text. Interestingly, the percentage of *uh(m)*'s in the source text does not yield a significant effect, so the issue of the influence of the *uh(m)*'s in the source text remains open. At the 10% significance level, there are two other significant predictors: the percentage of numerals in both the source text and the target text. They exhibit opposite effects, however. The numerals in the source text show a positive trend, which accords to expectation: if the source text contains more numerals, the interpreter has

**Table 5.** Parameter estimates for Analysis 2

| | Estimate | Std. Error | *t* value | Pr(>\|t\|) |
|---|---|---|---|---|
| Intercept | −2.987 | 0.066 | −45.563 | <0.001 |
| Delivery rate of SOURCE | 0.243 | 0.086 | 2.834 | 0.006 |
| Lexical density of SOURCE | 0.018 | 0.093 | 0.195 | 0.846 |
| % Numerals of SOURCE | 0.211 | 0.121 | 1.743 | 0.084 |
| Avg. Sentence length of SOURCE | 0.005 | 0.089 | 0.052 | 0.959 |
| % *Uh(m)'s* of SOURCE | 0.050 | 0.071 | 0.703 | 0.483 |
| Delivery rate of TARGET | −0.137 | 0.099 | −1.382 | 0.170 |
| Lexical density of TARGET | 0.041 | 0.091 | 0.456 | 0.650 |
| % Numerals of TARGET | −0.275 | 0.144 | −1.911 | 0.059 |
| Avg. Sentence length of TARGET | 0.087 | 0.096 | 0.902 | 0.370 |

significantly more difficulty with it and produces more *uh(m)*'s. By contrast, the numerals in the target text reveal a negative trend: if there are more numbers in the target text, the interpreter tends to produce fewer *uh(m)*'s. This result has also showed up in Analysis 1, so the effect is consistent. A tentative explanation for the discrepancy between the numerals in the source text and target text may be *omission*: if the numerals in the source text are positively correlated with the *uh(m)*'s in the target text, and the numerals in the target text are negatively correlated with it, then it is likely that some numbers are not translated. If that is indeed the case, then further research has to settle this issue as well as answer the question to which numbers the omission exactly pertains.

## 6. Discussion

In two cases we were able to identify determinants triggering significantly higher frequencies of *uh(m)*'s in the interpretations: a high source text delivery rate and a high target text lexical density. The results are inconclusive with regard to the proportion of numerals. This finding may be suggestive of different coping tactics by different interpreters, but such an explanation is difficult to pin down given the coarse-grained approach that we adopted. Nevertheless, our results lead to some interesting insights. First, different determinants are identified as triggers on the input side and on the output side of interpretation. This observation implies that different types of tasks cause cognitive overload in the listening effort and in the production effort. High delivery rates are unproblematic in production for both interpreters and speakers, but interpreters faced with high incoming delivery rates experience cognitive overload. In contrast, high lexical density is unproblematic in the listening effort, but challenges the production effort in interpreters. This asymmetry lends empirical support to theories of differential efforts (Gile 1995; Seeber 2011, 2013), as the listening effort and the production effort exceed the available cognitive resources on different tasks. Second, this study lends indirect support for the hypothesis that different efforts compete for the same cognitive resources. Indeed, if we assume that interpreters are not likely to produce denser texts than spontaneous speakers, the pattern emerging from the lexical density effects looks as follows: interpreters can cope with incoming texts that are lexically dense, as long as they do not try to produce texts that are lexically dense too. In other words, intensified lexical access in the listening effort, requiring high levels of cognitive resources, does not hinder lexical access at ordinary rates in production, but it does seem to prevent intensification of the productive lexical access. The production effort can thus be said to suffer from the allocation of more resources to the listening effort.

## 7. Conclusion

This article attempted to measure the cognitive or informational load in interpreting by modelling the occurrence rate of the disfluency *uh(m)*. In a corpus of 107 interpreted and 240 non-interpreted texts, informational load was operationalized in terms of the four measures of delivery rate, lexical density, percentage of numerals and average sentence length. The analysis proceeded in two ways. First, the 240 non-interpreted texts were compared with the 107 target texts in order to uncover the difference in production effort between interpreting and spontaneous speech. Secondly, the measures in the 107 target texts were compared to the measures in the 107 source texts, together with the percentage of source-*uh(m)*'s, in order to capture the listening effort next to the production effort in interpreting.

The results demonstrated that the interpreters produced significantly more *uh(m)*'s than non-interpreters and that this difference is mainly due to the effect of target text lexical density. That predictor exhibits a decreasing relation for the non-interpreters but the expected increasing relation in interpreting. The decreasing trend for the non-interpreters was attributed to the scripted and prepared nature of the parliamentary speeches. The effect of lexical density was no longer significant in the second analysis of the 107 source and target texts, where the main predictor of *uh(m)*'s in the target text was shown to be the delivery rate of the source text. On a more general level of 10% significance, the second analysis also revealed an increasing effect of the numerals in the source texts and a decreasing effect of the numerals in the target texts. That discrepancy led to speculations about omission, which is an issue to be explored in future research.

Most importantly, the results from this study show that the input side cannot be said to pose more challenges to the interpreters than the output side, even though interpreters are in control of the latter and not the former. The challenges which they face on both sides are of a different nature. This observation lends empirical support to interpretation models based on different efforts competing for cognitive resources.

Future avenues of research might include modeling the data at the level of individual sentences instead of at the level of entire texts. As that analysis is more fine-grained, the results may provide greater insight into the specific variables impacting interpreter performance. Furthermore, such an analysis will enable us to take a closer look at the position which the *uh(m)* occupies in the sequence of an utterance. It can be surmised that different sequential positions correlate with different degrees of cognitive load. In particular, we expect that there will be more *uh(m)*'s at the beginning of sentences in interpreting, whereas the *uh(m)*'s will be more evenly distributed in spontaneous speech.

## References


Agresti, Alan. 2013. *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons.

Alessandrini, Maria S. 1990. "Translating Numbers in Consecutive Interpretation: An Experimental Study." *The Interpreters' Newsletter* 3: 77–80.

Arnold, Jennifer E., Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. "Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering." *Language* 76 (1): 28–55. doi:10.1353/lan.2000.0045

Arnold, Jennifer E., Maria Fagnano, and Michael K. Tanenhaus. 2003. "Disfluencies Signal Theee, Um, New Information." *Journal of Psycholinguistic Research* 2003 (3): 25–36. doi:10.1023/A:1021980931292

Bachy, Sylviane, Anne Dister, Michel. Francard, Geneviève Geron, Vincent Giroul, Philippe Hambye, Anne-Catherine Simon, and Régine Wilmet. 2007. *Conventions de Transcription Régissant les Corpus de la Banque de Données VALIBEL*. [Transcription Conventions of the Corpora Included in the VALIBEL Database] <www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.PDF>

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology: In Honour of John Sinclair*, ed. by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233–250. Philadelphia: John Benjamins. doi:10.1075/z.64.15bak

Barik, Henri C. 1975. "Simultaneous Interpretation: Qualitative and Linguistic Data." *Language and Speech* 18 (3): 272–297.

Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober, and Susan E. Brennan. 2001. "Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender." *Language and Speech* 44 (2): 123–147. doi:10.1177/00238309010440020101

Broadbent, Donald E. 1958. *Perception and Communication*. London: Pergamon Press. doi:10.1037/10037-000

Cecot, Michela. 2001. "Pauses in Simultaneous Interpretation: A Contrastive Analysis of Professional Interpreters' Performances." *The Interpreters' Newsletter* 11: 63–85.

Chmiel, Agnieszka, and Iwona Mazur. 2013. "Eye Tracking Sight Translation Performed by Trainee Interpreters." In *Tracks and Treks in Translation Studies*, ed. by Way, Catherine, Sonia Vandepitte, Reine Meylaerts, and Magdalena Bartłomiejczyk, 189–205. Philadelphia: John Benjamins. doi:10.1075/btl.108.10chm

Clark, Herbert H., and Jean E. Fox Tree. 2002. "Using *uh* and *um* in Spontaneous Speaking." *Cognition* 84: 73–111. doi:10.1016/S0010-0277(02)00017-3

Dillinger, Mike. 1994. "Comprehension during Interpreting: What do Interpreters know that Bilinguals Don't?" In *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, ed. by Sylvie Lambert and Barbara Moser-Mercer, 155–189. Philadelphia: John Benjamins. doi:10.1075/btl.3.14dil

Faraway, Julian J. 2006. *Extending the Linear Model with R*. Boca Raton: Chapman & Hall/CRC.

Flesch, Rudolph. 1948. "A New Readability Yardstick." *Journal of Applied Psychology* 32 (3): 221–233. doi:10.1037/h0057532

Fox, John. 2003. "Effect Displays in R for Generalised Linear Models." *Journal of Statistical Software* 8 (15): 1–27. doi:10.18637/jss.v008.i15

Fox, John, and Sanford Weisberg. 2011. *An R Companion to Applied Regression*, Second Edition. Thousand Oaks, London and New Delhi: SAGE Publications.

Gerver, David. 1969. "The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters." In *Proceedings of the 2nd Louisville Conference on Rate and/or Frequency Controlled Speech*, ed. by Emerson Foulke, 162–184. University of Louisville: Centre for Rate-Controlled Recordings.

Gerver, David. 1975. "A Psychological Approach to Simultaneous Interpretation." *Meta* 20 (2): 119–128. doi:10.7202/002885ar

Gerver, David. 1976. "Empirical Studies of Simultaneous Interpretation: A Review and a Model." In *Translation: Applications and Research*, ed. by Richard W. Brislin, 165–207. New York: Gardner Press.

Gibson, Tim. R. 1993. *Towards a Discourse Theory of Abstracts and Abstracting*. Nottingham: University of Nottingham.

Gile, Daniel. 1995. *Regards sur la recherche en interprétation de conférence*. Lille: Presses Universitaires de Lille.

Gile, Daniel. 1997. "Conference Interpreting as a Cognitive Management Problem." In *Cognitive Processes in Translation and Interpreting*, ed. by Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain, and Michael McBeath, 196–214. Thousand Oaks, CA: SAGE Publications.

Gile, Daniel. 1999. "Testing the Effort Model's Tightrope Hypothesis in Simultaneous Interpreting – A Contribution." *Hermes* 22: 51–79.

Gile, Daniel. 2008. "Local Cognitive Load in Simultaneous Interpreting and its Implications for Empirical Research." *Forum* 6: 59–77. doi:10.1075/forum.6.2.04gil

Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins. doi:10.1075/btl.8

Goldman-Eisler, Frieda. 1967. "Sequential Temporal Patterns and Cognitive Processes in Speech." *Language and Speech* 10 (3): 122–132.

Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, and Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. [General Dutch Grammar.] Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.

Kintsch, Walter, Ely Kozminsky, William J. Streby, Gail McKoon, and Janice M. Keenan. 1975. "Comprehension and Recall of Text as a Function of Content Variables." *Journal of Verbal Learning and Verbal Behavior* 14 (2): 158–169. doi:10.1016/S0022-5371(75)80065-X

Levelt, Willem. 1983. "Monitoring and Self-Repair in Speech." *Cognition* 14: 41–104. doi:10.1016/0010-0277(83)90026-4

Mazza, Cristina. 2001. "Numbers in Simultaneous Interpretation." *The Interpreters' Newsletter* 11: 87–104.

Mead, Peter. 2000. "Control of Pauses by Trainee Interpreters in Their A and B Languages." *The Interpreters' Newsletter* 10: 89–102.

Mizuno, Akira. 1999. "Shifts of Cohesion and Coherence in Simultaneous Interpretation from English into Japanese." *Interpreting Research* 8 (2): 31–41.

Moser-Mercer, Barbara. 1978. "Simultaneous Interpretation: A Hypothetical Model and its Practical Application." In *Language interpretation and communication. Proceedings of the NATO symposium, Venice, Italy, September 26–October 1*, 1977, ed. by David Gerver and Wallace Sinaiko, 353–368. New York and London: Plenum Press.

Oostdijk, Nelleke. 2000. "The Spoken Dutch Corpus: Overview and First Evaluation." In *Proceedings of the Second International Conference on Language Resources and Evaluation*, ed. by Maria Gravilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis, and Gregory Stainhaouer, 887–894. Paris: ELRA.

Paradis, Michel. 1994. "Toward a Neurolinguistic Theory of Simultaneous Translation: The Framework." *International journal of psycholinguistics* 9 (3): 319–335.

Pinochi, Diletta. 2009. "Simultaneous Interpretation of Numbers: Comparing German and English to Italian. An Experimental Study." *The Interpreters' Newsletter* 14: 33–57.

Pio, Sonia. 2003. "The Relation Between St Delivery Rate and Quality in Simultaneous Interpretation." *The Interpreters' Newsletter* 12: 69–100.

Riccardi, Alessandra. 1998. "Interpreting Strategies and Creativity." In *Translator's Strategies and Creativity*, ed. by Ann Beylard-Ozeroff, Jana Kralova, and Barbara Moser-Mercer. Philadelphia: John Benjamins. doi:10.1075/btl.27.24ric

Seeber, Kilian. 2011. "Cognitive Load in Simultaneous Interpreting: Existing Theories – New Models." *Interpreting* 13 (2): 176–204. doi:10.1075/intp.13.2.02see

Seeber, Kilian. 2013. "Cognitive Load in Simultaneous Interpreting: Measures and Methods." *Target* 25 (1): 18–35. doi:10.1075/target.25.1.03see

Setton, Robin. 1999. *Simultaneous Interpretation: A Cognitive-pragmatic Analysis*. Philadelphia: John Benjamins. doi:10.1075/btl.28

Shlesinger, Miriam. 1995. "Shifts in Cohesion in Simultaneous Interpreting." *The Translator* 1 (2): 193–214. doi:10.1080/13556509.1995.10798957

Shlesinger, Miriam. 1998. "Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies." *Meta* 43 (4): 486–493. doi:10.7202/004136ar

Shlesinger, Miriam. 2003. "Effects of Presentation Rate on Working Memory in Simultaneous Interpreting." *The Interpreter's Newsletter* 12: 37–49.

Straniero Sergio, Francesco, and Caterina Falbo. 2012. *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang. doi:10.3726/978-3-0351-0377-9

Swerts, Marc. 1998. "Filled Pauses as Markers of Discourse Structure." *Journal of Pragmatics* 30: 485–496. doi:10.1016/S0378-2166(98)00014-9

Tissi, Benedetta. 2000. "Silent Pauses and Disfluencies in Simultaneous Interpretation: A Descriptive Analyzes." *The Interpreters' Newsletter* 10: 103–127.

Tommola, Jorma, and Marketta Helevä. 1998. "Language Direction and Source Text Complexity: Effects on Trainee Performance in Simultaneous Interpreting." In *Unity in Diversity? Current Trends in Translation Studies*, ed. by Lynne Bowker, Michael Cronin, Dorothy Kenny, and Jennifer Pearson, 177–186. Manchester: St. Jerome.

Van de Kauter, Marjan, Geert Coorman, Els Lefever, Bart Desmet, Lieve Macken, and Veronique Hoste. 2013. "Lets Preprocess: The Multilingual Lt3 Linguistic Preprocessing Toolkit." *Computational Linguistics in the Netherlands journal* 3: 103–120.

Voor, John B., and Joseph M. Miller. 1965. "The Effect of Practice on the Comprehension of Worded Speech." *Speech Monographs* 32: 452–455. doi:10.1080/03637756509375469

Watanabe, Michiko, Keikichi Hirose, Yasuharu Den, and Nobuaki Minematsu. 2008. "Filled Pauses as Cues to the Complexity of Up-Coming Phrases for Native and Non-Native Listeners." *Speech Communication* 50 (2): 81–94. doi:10.1016/j.specom.2007.06.002

Welford, Alan T. 1952. "The 'Psychological Refractory Period' and the Timing of High Speed Performance — A Review and a Theory." *British Journal of Psychology* 43: 2–19.

## Appendix

**Table A.1.** Correlation matrix of the variables in Analysis 1

| | **log(Lex. dens.)** | **log(% Num.)** | **log(Avg. sen. len.)** |
|---|---|---|---|
| log(Del. rate) | −0.372 | −0.220 | 0.171 |
| log(Lex. dens.) | | 0.183 | 0.054 |
| log(% Num.) | | | −0.194 |

**Table A.2.** Correlation matrix of the variables in Analysis 2

| | **log(SRC Lex. dens.)** | **log(SRC % Num.)** | **log(SRC Avg. sen. len.)** | **log(SRC % *Uh(m)*)** | **log(TGT Del. Rate)** | **log(TGT Lex. dens.)** | **log(TGT % Num.)** | **log(TGT Avg. sen. len.)** |
|---|---|---|---|---|---|---|---|---|
| log(SRC Del. rate) | −0.047 | −0.183 | −0.122 | 0.064 | 0.550 | −0.020 | −0.038 | −0.241 |
| log(SRC Lex. dens.) | | 0.118 | −0.034 | −0.425 | 0.059 | 0.525 | 0.053 | 0.151 |
| log(SRC % Num.) | | | −0.028 | −0.229 | −0.269 | 0.141 | 0.858 | −0.099 |
| log(SRC Avg. sen. len.) | | | | 0.161 | −0.266 | −0.127 | −0.113 | 0.631 |
| log(SRC % *Uh(m)*) | | | | | −0.026 | −0.253 | −0.160 | 0.050 |
| log(TGT Del. Rate) | | | | | | −0.156 | −0.258 | −0.014 |
| log(TGT Lex. dens.) | | | | | | | 0.143 | −0.182 |
| log(TGT % Num.) | | | | | | | | −0.230 |

*Authors' addresses*

Koen Plevoets
EQTIS, Ghent University
Groot-Brittanniëlaan 45
B 9000 Gent, Belgium

koen.plevoets@ugent.be

Bart Defrancq
EQTIS, Ghent University
Groot-Brittanniëlaan 45
B 9000 Gent, Belgium

bart.defrancq@ugent.be